

# Numpy. Broadcasting. Векторизация

Гончаров Павел [kaliostogoblin3@gmail.com](mailto:kaliostogoblin3@gmail.com)  
Нестереня Игорь [nesterione@gmail.com](mailto:nesterione@gmail.com)

# Numpy



NumPy

[Scipy.org](http://www.numpy.org)

## NumPy

NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional array object
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortran code
- useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

NumPy is licensed under the [BSD license](#), enabling reuse with few restrictions.

## Getting Started

- [Getting NumPy](#)
- [Installing the SciPy Stack](#)
- [NumPy and SciPy documentation page](#)
- [NumPy Tutorial](#)
- [NumPy for MATLAB® Users](#)
- [NumPy functions by category](#)
- [NumPy Mailing List](#)

For more information on the SciPy Stack (for which NumPy provides the fundamental array data structure), see [scipy.org](http://www.scipy.org).

Модуль numerical python (Numpy <http://www.numpy.org>) для работы с матрицами. Векторизация вместо циклов. Индексация и срезы в массивах. Чтение и запись файлов с помощью Numpy.

# Установка SciPy

```
pip3 install numpy scipy matplotlib ipython jupyter pandas sympy nose
```

Для windows проще всего взять готовые пакеты:

<http://www.lfd.uci.edu/~gohlke/pythonlibs/>

# Создание вектора/матрицы

```
import numpy as np
```

```
a = np.array([1,2,3,4])
```

```
b = np.array([[1,2,3],[4,5,6]])
```

```
E = np.eye(5)
```

```
d = np.diag([1,2,3,4])
```

```
z=np.zeros((2,3))
```

# Операции

Numpy массивы - объект, для которого переопределены многие операции и он содержит методы для работы с матрицами.

Некоторые из них:

a.T - транспонирование

a.sum - вернуть сумму элементов

a.shape - размерность

# Broadcasting

Способность NumPy работать с массивами разных размерностей во время арифметических операций. Операции над массивами обычно выполняются для соответствующего элемента.

# Векторизация

Замена циклов в коде операциями над матрицами векторами.

Пример вычисления MSE

$$MSE = \frac{1}{m} \sum_i (y^{pred} - y)^2 = \frac{1}{m} \|y^{pred} - y\|_2^2 = \frac{1}{m} (y^{pred} - y)(y^{pred} - y)^T$$

```
def mse(y_pred, y):  
    return (y_pred - y).T.dot(y_pred - y) / len(y_pred)
```

# Постановка задачи машинного обучения



# Формальное определение

A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ .” (Mitchell 1997)

Примеры:

$T$  - регрессия, классификация, ...

$E$  - примеры для обучения (для обучения с учителем)

$P$  - MSE, RMSE...

# Регрессия

задача регрессии построить зависимость между вектором  $x$  и некоторым значением  $y$ .

Например:

Предсказание цены дома

<https://www.kaggle.com/c/house-prices-advanced-regression-techniques>

# Часть набора данных для обучения

<u>OpenPorchSF</u>	<u>EnclosedPorch</u>	<u>3SsnPorch</u>	<u>ScreenPorch</u>	<u>PoolArea</u>	<u>PoolQC</u>	<u>Fence</u>	<u>MiscFeature</u>	<u>MiscVal</u>	<u>MoSold</u>	<u>YrSold</u>	<u>SaleType</u>	<u>SaleCondition</u>	<u>SalePrice</u>
61	0	0	0	0	NA	NA	NA	0	2	2008	WD	Normal	208500
0	0	0	0	0	NA	NA	NA	0	5	2007	WD	Normal	181500
42	0	0	0	0	NA	NA	NA	0	9	2008	WD	Normal	223500
35	272	0	0	0	NA	NA	NA	0	2	2006	WD	Abnorml	140000
84	0	0	0	0	NA	NA	NA	0	12	2008	WD	Normal	250000
30	0	320	0	0	NA	MnPrv	Shed	700	10	2009	WD	Normal	143000
57	0	0	0	0	NA	NA	NA	0	8	2007	WD	Normal	307000
204	228	0	0	0	NA	NA	Shed	350	11	2009	WD	Normal	200000
0	205	0	0	0	NA	NA	NA	0	4	2008	WD	Abnorml	129900
4	0	0	0	0	NA	NA	NA	0	1	2008	WD	Normal	118000
0	0	0	0	0	NA	NA	NA	0	2	2008	WD	Normal	129500
21	0	0	0	0	NA	NA	NA	0	7	2006	New	Partial	345000
0	0	0	176	0	NA	NA	NA	0	9	2008	WD	Normal	144000
33	0	0	0	0	NA	NA	NA	0	8	2007	New	Partial	279500
213	176	0	0	0	NA	GdWo	NA	0	5	2008	WD	Normal	157000
112	0	0	0	0	NA	GdPrv	NA	0	7	2007	WD	Normal	132000
0	0	0	0	0	NA	NA	Shed	700	3	2010	WD	Normal	149000

# Задача машинного обучения

## обобщение (generalization)

Задача минимизировать ошибку на тестовых данных (не на тех, которых производилось обучение)

Цель алгоритма предсказывать значения на новых данных (из того же распределения)

## Линейная регрессия

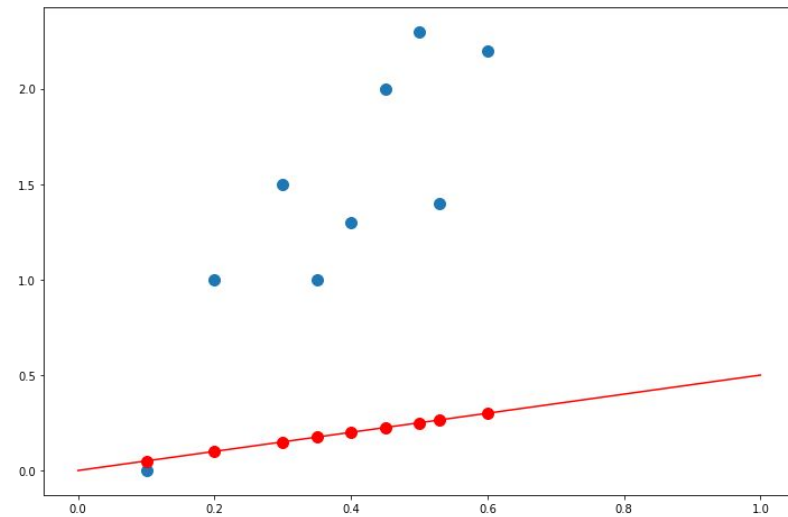
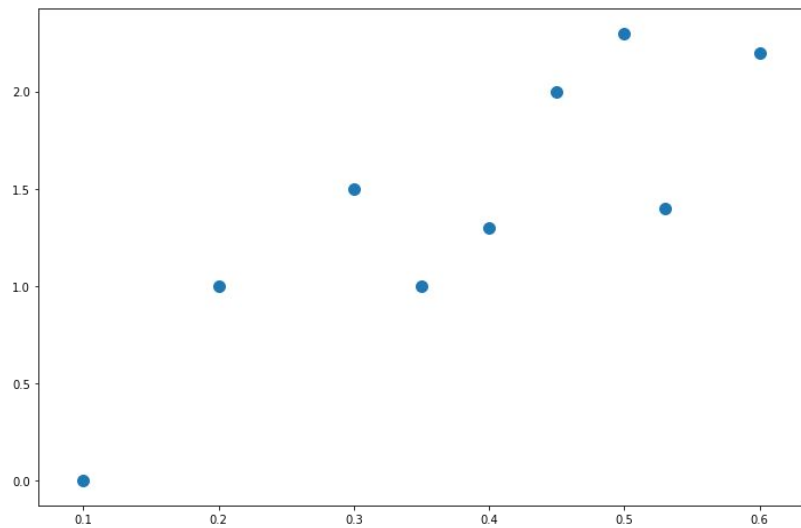
$$y^{pred} = w^T x$$

Для задачи регрессии в качестве оценки производительности используется средняя квадратичная ошибка (mean squared error) (MSE):

$$MSE = \frac{1}{m} \sum_i (y^{pred} - y)^2$$

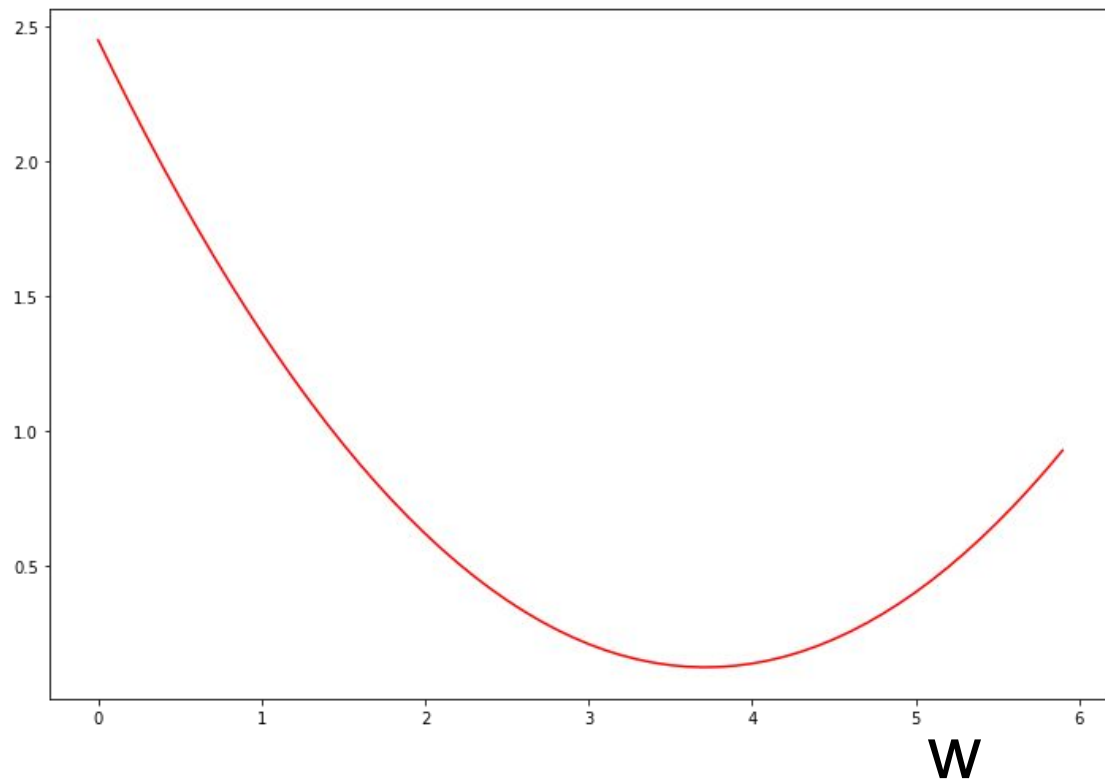
или можно записать через норму:

$$MSE = \frac{1}{m} \|y^{pred} - y\|_2^2$$



MSE = 1.8641083333

MSE





# Обучение без обучения

Для данной задачи можно решить аналитически.

$$\nabla_{\mathbf{w}} \text{MSE}_{\text{train}} = 0$$

<http://www.deeplearningbook.org/> (106-107)

MSE = 0.15

